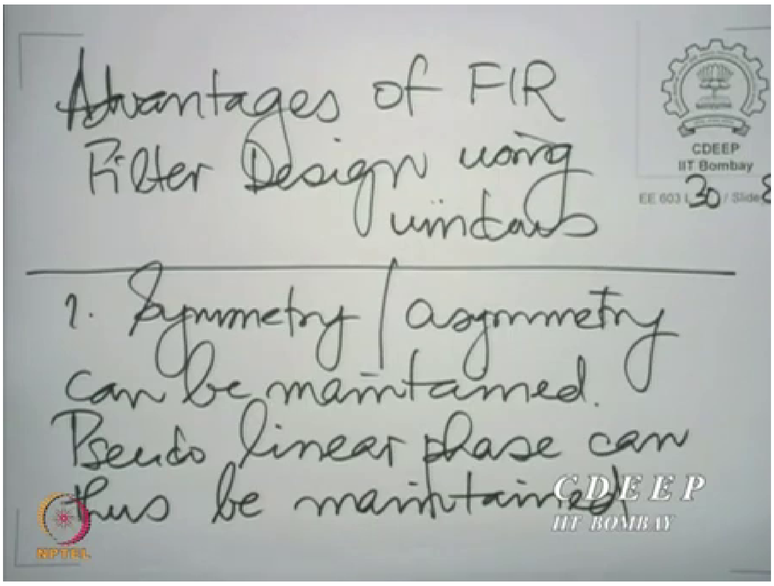
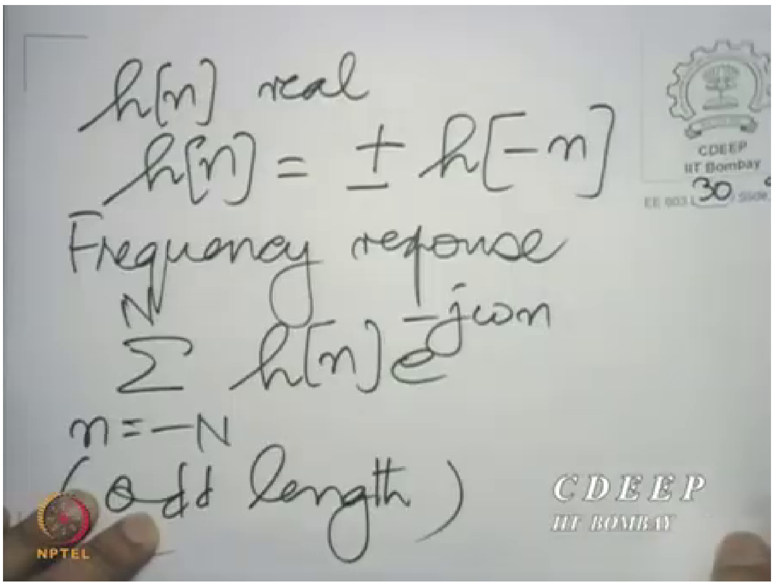
Digital Signal Processing and its Applications
Professor Vikram M. Gadre
Department of Electrical Engineering
Indian Institute of Technology, Bombay
Lecture 30 B
Advantages of FIR Design Using Windows

(Refer Slide Time: 00:20)



So you see, what we need to do is to put down some advantages of FIR filter design by window. First symmetry or asymmetry can be maintained and therefore, linear phase or pseudo linear phase can therefore be maintained as well. Let me spend a minute explaining this once again. It is a very important idea.

(Refer Slide Time: 01:29)

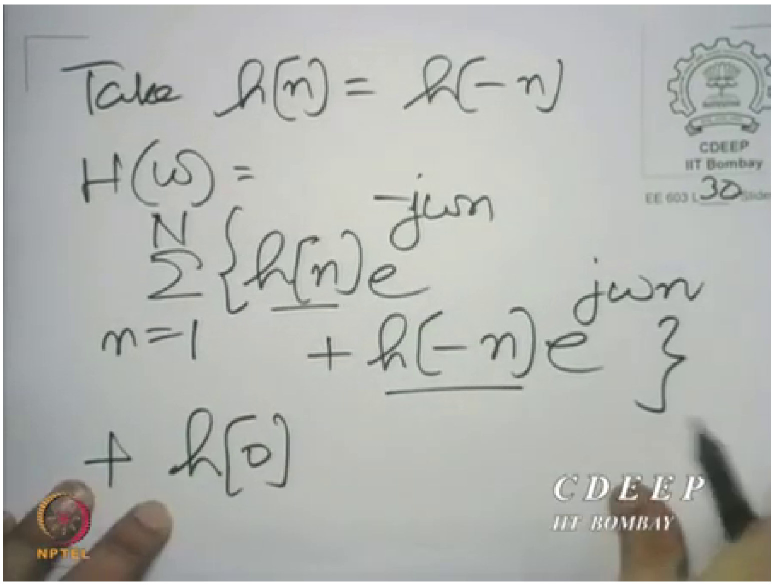


What you are saying is that, if you have a response h[n], which is real and

$$h[n] = \pm h[-n]$$

I am sorry, there is either even symmetry or odd symmetry then, the corresponding frequency response is of the form summation n going from $-N$ to $+N$, now here, I am assuming odd length, the same argument can be extended to even length h[n] $e^{-j\omega n}$ and we can club, we will take the plus and minus cases separately.

(Refer Slide Time: 02:36)

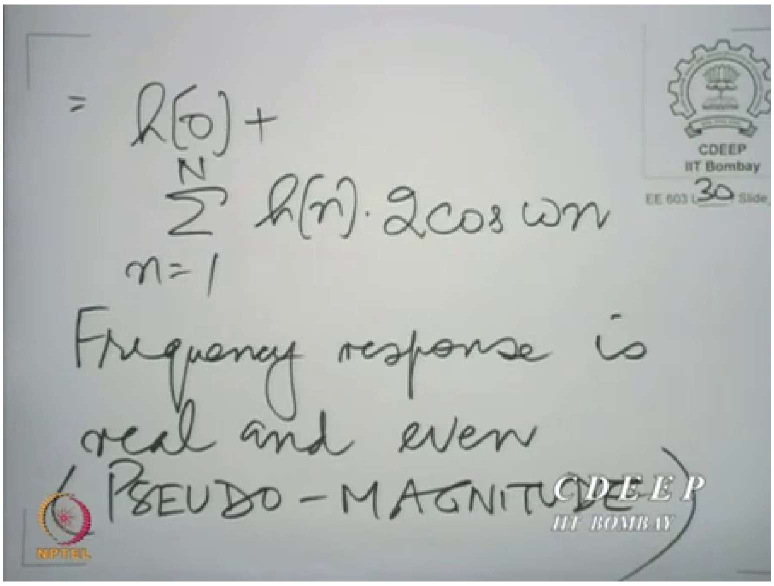


So, we can club to take $h[n] = h[-n]$. So we can rewrite this as

$$H(w) = \sum_{n=1}^{N} h[n] \cdot e^{-j\omega n} + h[-n] \cdot e^{j\omega n} + h[0],$$

by clubbing the plus and minus terms together and these are equal.

(Refer Slide Time: 03:28)



And therefore we can rewrite this as,

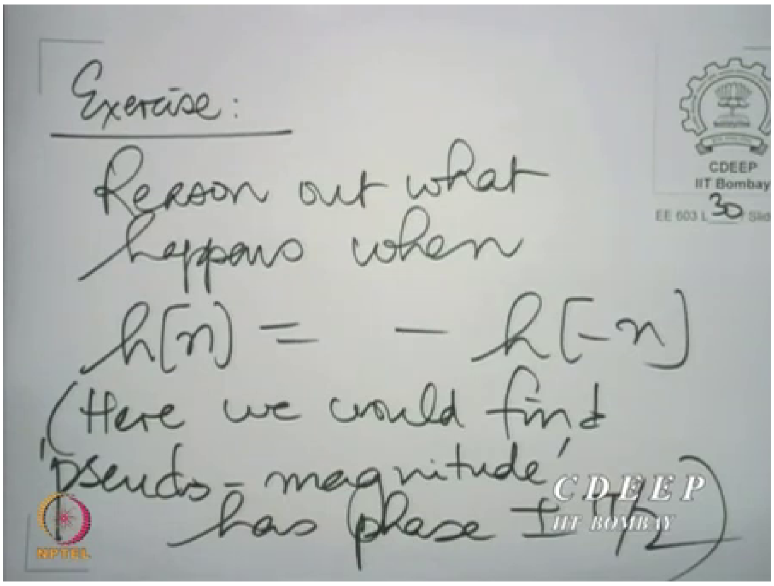
$$= h[0] = \sum_{n=1}^{N} h[n] \cdot 2cos\omega n$$

So, therefore when you have a real and even impulse response, as we expect the frequency response is also real and even and this is called the pseudo magnitude. It is called a pseudo magnitude because if you now delay this impulse response by n samples to make the FIR filter causal, the only change that takes place in the frequency response is a factor of $e^{-j\omega}$ times the delay.

And that only contributes a linear phase. So, you have a pseudo magnitude multiplied by a linear phase. Now, the only catch is that it is a pseudo magnitude. This is not quite the magnitude. In other words, it could be positive or negative, wherever it is negative; you are also putting in additional phase of π (pi).

So, you can call this resultant causal FIR filter, that is the FIR filter which has been obtained by delaying the spike capital n samples as a pseudo linear phase filter. Pseudo in the sense that, it is linear phase to the extent of a, to a phase factor of pi; linear phase plus minus, I mean plus 0 or +π. It is called pseudo linear phase and a similar, in fact, this I leave to you as an exercise

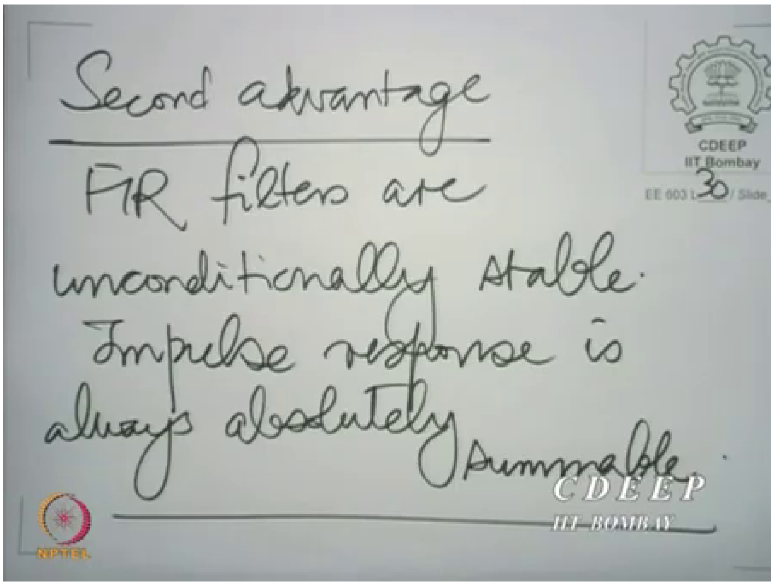
(Refer Slide Time: 05:31)



Exercise: Reason out, what happens when $h[n] =- h[- n]$. Here you would find the pseudo magnitude has phase $\pm \pi/2$. Essentially, what we call the pseudo magnitude in this case would have an additional factor of either +j or -j and then if you delay it again, you have the linear phase term, but then here are the pseudo, so called pseudo magnitude would have a phase of either +90° or -90°.

So, this is what we mean by FIR filters, allowing us linear phase. When you maintain symmetry or anti symmetry in the response, then you are guaranteed a pseudo magnitude or pseudo linear phase, pseudo magnitude plus linear phase or pseudo linear phase. So, it is the best, the closest to linear phase that we can get. That is what it means.

(Refer Slide Time: 07:02)



Now, this is one of the advantages. The second advantage is that FIR filters are unconditionally stable, impulse responses is absolutely summable, even in the presence of numerical inaccuracies. So you see if the coefficients are real, when we realize the coefficients in finite precision that is likely to be inaccuracy in the representation of the coefficients. But even in the presence of those inaccuracies, the stability of the filter is unaffected.

Now, this is not the case with IIR filters. If the poles of IIR filter happened to be close to the unit circle and if there are numerical inaccuracies in realizing the coefficients, there is a possibility that the pose may migrate outside the unit circle in the presence of numerical inaccuracies and then we have trouble instability.

Then it does not remain a filter at all. Because then you are not even, you are not sure, if the now, I would still say it remains maybe it is not correct to say that it does not remain a filter, it remains a filter, but then you have this trouble that you are not sure whether a bounded input can result in a bounded output or not.

Incidentally, IIR filters can never give you linear phase. In fact, I poses a challenge to you shows that IIR filters, which are causal can never give you a linear phase. I believe I posed this challenge before but I am just repeating the challenge again. I also give you a hint, the heat lies in showing that causality and symmetric cannot go together; causality, symmetry and IIR cannot all go together.

Anyway, you see we have seen this universal principle of engineering and nothing comes for free. This is also true here. So FIR filters seem to have everything that we would want them to; in fact, one more thing that they have is that there is at least a design approach for FIR filters which are non-piecewise constant in the ideal response.

So we know how to, for example, realize an approximation to the discrete time differentiator by using FIR filter; simplifying the ideal impulse response and truncate it or find the ideal impulse response and then window it. So we know I mean at least one way to do it, we do not know how well that approach would work.

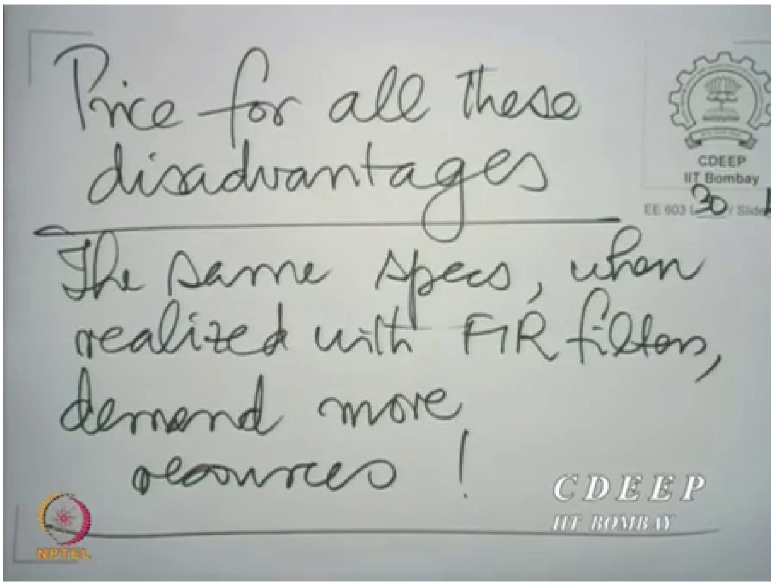
But experience tells us that it at least works well we have an approach. We do not have one for IIR features at all. There is no way to design an IIR or there is no easy way to design an IIR discreetly differentiate, I mean not, definitely not based on the bilinear transform, because a bilinear transform is going to distort the frequency axis.

So, it cannot be used for non-piecewise constant responses. You see the, if now when you reflect on the bilinear transform with the benefit of hindsight, you realize that the reason why the bilinear transform worked even though it made a nonlinear distortion of the frequency axis is that the bilinear transform in frequency was a monotonically increasing transfer. So as capital omega increase small omega also increased all the way from minus to plus infinity.

So, as capital omega the analog frequency went from minus infinity to plus infinity, the discrete time frequency went from $-\pi$ to π and therefore, pieces of the axis contiguous pieces of the frequency axis map to corresponding similarly ordered contiguous pieces of the discrete frequency axis, pieces went to pieces and therefore, the bilinear transform in spite of the non-linearity of the frequency transformation was improbable for piecewise constant filter design.

But it would not be applicable for designing a discrete time differentiator. Because there, even if you happen to design a very good analog band limited differentiator, when you transform it with the bilinear transform, the frequency response will be completely distorted from linear and therefore, we do not have a good, right now we do not have, we have not talked about any meaningful way, design, discrete differentiators or similar such responses which are not piecewise constant in the IIR context. That is another reason again, why IIR, why FIR filters are attractive. So then, where is the, where is the price that we paying? The price and that is what we will now write.

(Refer Slide Time: 12:32)

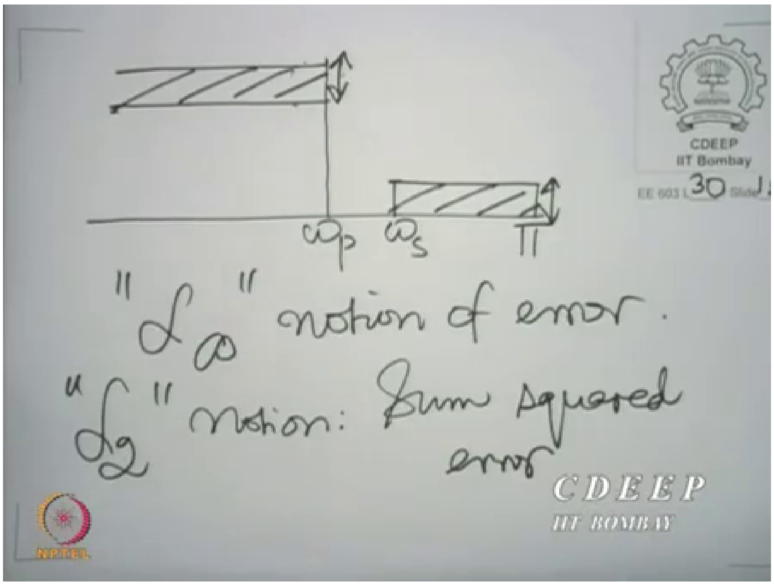


Price for all these advantages; the same specifications when realized with FIR filter designs, demand more resources. So, you would want to verify this when you carry out the design that you have been assigned. For the same magnitude specifications, when we realize it using the FIR filter, you would find typically that the FIR filter is much longer; it requires many more additions and derails.

So nothing comes for free. Anyway, so it is all about the relative behaviour of IIR and FIR filters and now we have been talking about resources all this while, we must now actually come down to the issue of realizing filters. Now, there is one little thing before I go to realization that I would like to mention in the context of FIR filter design. You see, one might wonder why at all one should use the rectangular window when you have so many other windows to choose from?

Of course one argument is that the transition bandwidth is kind of the minimum. So transition bandwidth is the issue, then the rectangular window is a good choice. But more importantly, there is a fundamental other reason why the rectangular window is attractive; you see, when we talk about pass band and stop band tolerance all this while what we have been talking about is, what is called the L_∞ tolerance or the maxim deviation.

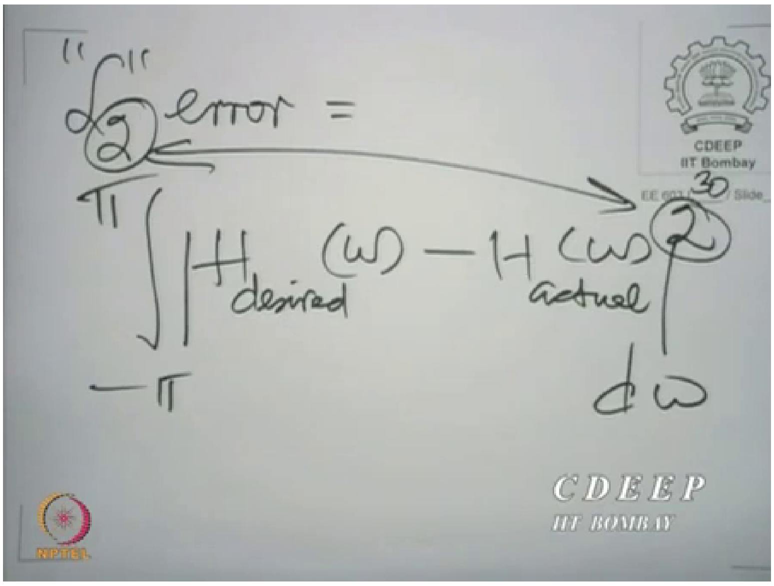
(Refer Slide Time: 14:36)



So what I am saying is, when you put down the specs for a low pass filter, for example, we said something like this, we said there is a passband tolerance and there is stop band tolerance meaning that the magnitude in the passband must be within this shaded area and the magnitude in the stop band must be within the shaded area. However, we are saying nothing at all about the extent to which or how or the frequency with which this should deviate from the ideal in the pass band and the stop band.

So it is quite possible that in the pass band, it is only at one frequency that it really goes all the way up to the tolerance everywhere else, it might be far away from the tolerance, it might be close to the ideal. So this is called the L_∞ notion of error. Now, this L_∞ is a strange word at the moment, but it will become clearer when we come to another notion of error, called the L_2 notion of error. The L_2 notion would be the mean squared error or the sum squared error and it will be very clear where the number 2 comes from.

(Refer Slide time: 16:06)



You see that the sum squared error, L_2 error; as you might want to call it is essentially, the desired frequency response minus the actual frequency response. The absolute value this is taken and integrated from $-\pi$ to π and this is where the 2 comes from. The 2 comes from the square. So, when we talk about L_2 error, what we are really talking about is the magnitude squared of the error has seen all over the pass band.

You see now also the actually, if you really want to understand the number, the why it is called L_2, one should define the L_2 error to be the square root of this integral of the error square mod error squared. But it does not matter. I mean, you know, if the squared error is a maximum, so is the square root of it, the square root is a monotonically increasing function, that is not such an issue.

But if you do take the square root, then it does explain the infinity concept. So if you were to take this instead of 2, if you had 3 there, you would call it the L_3 error. If you had 1 there that means if you took just the absolute value and integrate, we call the L_1 error. So you could now conceive of the L_∞ error. That means you raise it to the power infinity notionally or raise it to a larger, larger power, but then do not forget to take one by that power outside.

So raise it to the corresponding root also. So if you are raising it to, if you take while calculating the L_{10} error, for example, you raise the mod error to the part 10. But then you take a $1/10$th power outside. Now you can visualize this being taken to infinity and then what really happens, is that as you take it to infinity, it is only the maximum which survives all the others are suppressed. That is why we call it… Yes, there is a question.

Student: (())(18:29)

Professor: So the question is, would you consider the error only over the pass band or everywhere? The answer is everywhere. You know, you have a desired response all over. Now, you may argue what happens in the transition band? Well, that is an important point, the transition band actually does not have an, have a desired response specified. Now, again, it is interesting, it does not matter.

So here, you could, for example, take the middle of the transition band as a point of separation and you could take the response to be one up to the middle and then zero after the middle in the case of a low pass filter, for example and use that as a desired response. You can also if you wish, take only the pass band and the stop band and put down error here. That will also be a meaningful L_2 error. But the error is calculated all over the band from 0 to π. Now, yes, there is a question?

Student: (())(19:40)

Professor: So the question is, how do you take the desired ω? The desired omega, the desired response is 1 in the pass band and 0 in the store band. Anyway, the point is the rectangular window actually minimizes the L_2 error as well. In addition to the transition band being optimized with the rectangular window, if one is talking, you see the transition band is optimized, but the L_∞ error is the worst for the rectangular window.

However, the L_2 error is the minimum. So, the rectangular window is not without advantages. So, you see that also tells us that L_2 error minimization is not the same as L_∞ error minimization and that is not too difficult to understand. You see, it is quite possible that at one place as I said, you may have the response deviating very far from the ideal, but it may be pretty close to the ideal many other places and therefore, the L_2 error could be low.

On average the squared error could be low, but because at one place it deviates very far, the L_∞ error is significant. So, that is about L_∞ and L_2 error. What I also tried to illustrate here is that there is just not one notion of error, though we have taken the L_∞ error all the time now discussion without having explicitly realized all this while.